Improving Ultrasound-Based Gesture Recognition by Partially Shielded Microphone

Hiroki Watanabe[1(✉)] and Tsutomu Terada[2]

[1] Hokkaido University, Kita 14, Nishi 9, Kita Sapporo 060-0814, Japan
hiroki.watanabe@ist.hokudai.ac.jp
[2] Kobe University, 1-1 Rokkodai, Nada Kobe 657-8501, Japan
tsutomu@eedept.kobe-u.ac.jp

Abstract. We propose a method to improve ultrasound-based in-air gesture recognition. Doppler effect is often used to recognize ultrasound-based gesture. However, increasing the number of gesture is difficult because of limited information obtained from that. In this study, we partially shield the microphone by a 3D printed cover. Acoustic characteristics of the microphone is changed by the cover, and it increases the obtained information. Since the proposed method utilizes a 3D printed cover and a pair of embedded speaker and microphone of a device, it does not require additional electrical device to improve gesture recognition. We implemented five microphone covers and investigated the performance of the proposed method in six gestures with four participants. Evaluation results confirmed that the recognition accuracy increased 12% in the most effective device by using the proposed method.

Keywords: Gesture recognition · Ultrasound · Mobile computing
Wearable computing

1 Introduction

Mobile/wearable computers are commonly available in these days, such as smartphone and smartwatch. We can intuitively control these devices by using touch screen. However, one of the major problem of a touch screen is an occlusion. When user interacts with the device by his/her finger, it hides the screen. Moreover, touch screen is needed to be directly touched. There are some situations that user does not want to/cannot touch the screen, for example, when user's hands are wet or dirty and when user wears gloves. Therefore, in-air gesture is useful for interaction with mobile/wearable devices. There are several methods to recognize in-air gesture. Additional device, such as infrared sensor [1], enables to recognize in-air gesture. However, this method needs additional electrical devices that require power supply. Although built-in camera can recognize in-air gesture [5], camera-based method is regarded as privacy-invasive. It also

K. Murao et al. (Eds.): MobiCASE 2018, LNICST 240, pp. 223–229, 2018.
https://doi.org/10.1007/978-3-319-90740-6_15

requires the user is within line of sight of cameras and high computational cost. Moreover, most smartwatches do not have built-in camera. Built-in speaker and microphone also can recognize in-air gesture [3,4]. The speaker transmits ultrasound, and the microphone receives the reflected ultrasound. These studies utilize Doppler effect or the time difference of arrival to the stereo microphone. However, Doppler effect can detect only whether the target is approaching or away and its velocity. Thus, recognizing the gesture direction is difficult. Moreover, smartwatch generally does not have multiple microphones.

In this study, we focus on the method that physically changes the characteristic of a sensor [2]. We apply this method to a microphone. Specifically, we attach a cover that changes the acoustic characteristics to the built-in microphone of the device. This method increases the features obtained from microphone and improves the gesture recognition. We proposed three types of microphone covers, and evaluation results confirmed that the recognition accuracy improved by 12% in the most effective device. The contributions of this paper are as follows: (1) We proposed a method to improve ultrasound-based gesture by using the microphone cover. (2) We designed three types of microphone covers and investigated that which is suitable for the proposed method. (3) Evaluation results confirmed that the microphone cover can improve ultrasound-based gesture recognition.

2 Proposed Method

2.1 Assumed Environment

In this study, the device transmits ultrasound from the speaker, and receives the reflected ultrasound by the microphone, as shown in Fig. 1. The system recognizes in-air gesture by computing the frequency shift caused by Doppler effect. As a characteristic of Doppler effect, received frequency becomes high during the sound source is approaching to the receiver and becomes low during the sound source is moving away from the receiver. In the assumed environment, a transmitter is the built-in speaker and a receiver is the built-in microphone. We set 20 kHz as the frequency of the transmitted ultrasound because most of human cannot hear around 20 kHz and existing smartphone/smartwatch can transmit it. Ultrasound is not adversely affected by the environmental sound and it is easy to separate from the environmental sound.

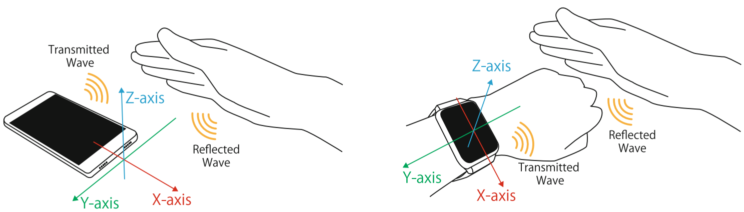


Fig. 1. Assumed environment

2.2 Cover Design

As reported in literature [4], the number of gesture is limited when we utilize Doppler effect with a pair of microphone and speaker. The left of Fig. 2 shows the gesture and correspond spectrogram. When the hand approaches to the microphone, frequency shift occurs in the high frequency side and when the hand moves away from the microphone, frequency shift occurs in the low frequency side. As this figure shows, it is difficult to classify left swipe and right swipe because spectrograms of both gestures are similar in this speaker-microphone configuration. The basic idea of the proposed method is changing acoustic characteristic of the microphone by just attaching microphone cover. As shown in the right of Fig. 2, we shield one side of the microphone, and change the sensitivity of the microphone between the left side and the right side. As the spectrogram of the right of Fig. 2 shows, the cover weaken frequency shift of the left side of the microphone and the difference between left swipe and right swipe is emphasized. Therefore, we consider that the system can classify these two gestures by the proposed method.

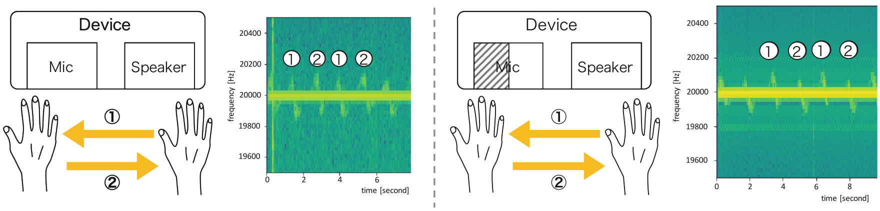


Fig. 2. Gestures and spectrograms of conventional method (left) and proposed method (right)

We consider three types of covers to partially shield the microphone as follows: opening multiple holes of 1 mm in diameter (Fig. 3A), closing with sponge that absorbs the ultrasound (Fig. 3B), and changing the directivity of the microphone (Fig. 3C). We preliminarily investigate which types of the shield is suitable for the proposed method. The used device was iPhone 5, and the target gestures are left swipe and right swipe. The participants are three. They performed each gesture 10 times. For simplicity, we just utilize the magnitude caused by the Doppler effect to classify gestures. When the high frequency magnitude caused by the Doppler effect is bigger than the lower one, we recognize the gesture is left swipe, and the gesture is recognized as right swipe in the opposite case. The result of preliminarily investigation is shown in the bottom of Fig. 3. As this figure shows, cover A and cover B had little effect. On the other hand, cover C could distinguish the left swipe and right swipe with high accuracy. Therefore, in this study, we adopt the design of cover C. Cover C has an angle of 45° to Y axis and Z axis to capture the gestures along Y/Z axis.

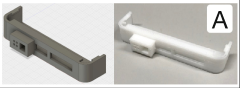
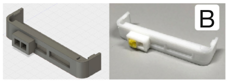
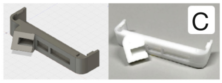
3D Model & Printed Cover	 A	 B	 C
Accuracy [%]	53.3	66.7	91.7

Fig. 3. Cover design candidates

2.3 Recognition Method

The recognition method is based on the method described in literature [4]. It includes the following step: (1) We compute the fast Fourier transform (FFT) to the input of the microphone, and since the input signal fluctuates, we normalize the FFT result of each frame. (2) To emphasize the magnitude change of frequency bins, we subtract the normalized spectrum of current frame by previous frame and then square the magnitudes of frequency bins. (3) To extract the reacted area, we utilize Gaussian smoothing to smooth, and binarize the smoothed data with a certain value.

In this study, we divide the area into two sections by the line of the frequency of the transmitted signal (20 kHz), as shown in Fig. 4. We compute 12 features from these two sections; width of upper/lower section, height of upper/lower section, area of upper/lower section, width ratio of upper section and lower section, height ratio of upper section and lower section, and area ratio of upper section and lower section. We utilize WEKA for classification and selected SVM for the classifier with default parameters.

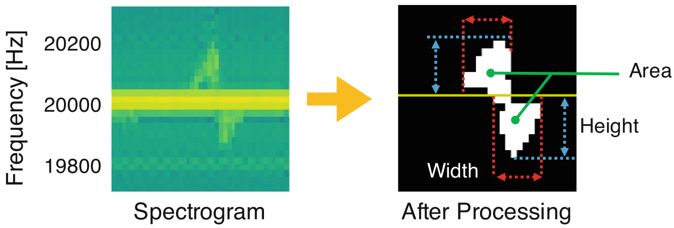


Fig. 4. Feature extraction

3 Implementation

We implemented the microphone covers based on the cover design discussed in Sect. 2.2. The implemented covers are shown in Fig. 5. The thickness of the cover was 1.5 mm. However, only the microphone part was 3 mm to enough shield ultrasound. The used devices were four smartphones (Apple iPhone 5, Apple iPhone 6s, Huawei P9 Lite, and Asus ZenFone 3 Laser) and one smartwatch (Asus ZenWatch 2). These all smartphones have a speaker at the bottom right side of the device and a microphone at the bottom left side of the device. Smartwatch has a speaker at the top of the left side and a microphone at the bottom of the left side. The used 3D printer was PP3DP UP Plus 2 and we used ABS filament to make the covers. We used Autodesk Fusion 360 to make 3D models.

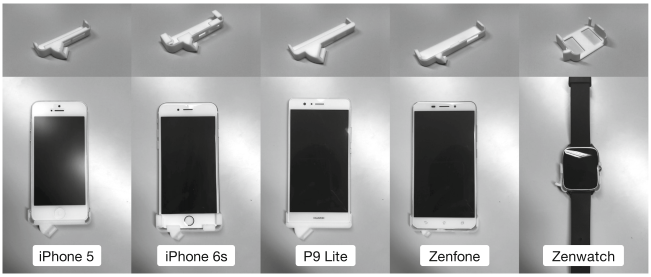


Fig. 5. Implemented cover

4 Evaluation

We evaluate the proposed method. In this study, we assume a situation that although the users want to browse recipe and manual during cooking and maintenance of the machine, they cannot touch the screen because their hand is wet or dirty. In this situation, we assume six gestures along three axes that is shown in Fig. 1. We assume that the gestures along X axis are scroll up/down, Y axis are go/back to the next/previous page, and Z axis are enter/cancel. The used devices are four smartphones and one smartwatch described in Sect. 3. The four participants are 22 to 27-year-old males and all of them use right hand to perform gesture. They conducted each six gestures 10 times. We evaluated the acquired data by 10-fold cross validation. When conducting the evaluation, the smartphones are put on the table and the smartwatch was worn on participants' left wrist, as shown in Fig. 1.

Table 1 shows the result of the evaluation. As this table shows, in most devices, the recognition accuracy improved when the device wears the microphone cover. Figure 6 shows the confusion matrix of iPhone 5 with/without cover. As this figure shows, the variation of the recognition result decreased in the whole. The effect of the cover differed depending on the devices. The most improvement was 12% at iPhone 5. On the other hand, the recognition accuracy of smartwatch decreased by attaching cover. This is because the detected volume of the reflected ultrasound was small even if the cover was not attached. Thus, the acquired volume became smaller when the cover attached, and we could not extract features effectively.

In this study, the gestures were performed with the devices placed in the center. However, if the device is not centered, the proposed method cannot recognize the gestures accurately because the characteristics of the gestures change in such situation.

Table 1. Recognition accuracy of evaluation with/without cover [%]

Participants	iPhone 5		iPhone 6s		P9 Lite		ZenFone		ZenWatch	
	w/o	w/	w/o	w/	w/o	w/	w/o	w/	w/o	w/
A	73.3	88.3	75	60	71.7	73.3	60	66.7	43.3	40
B	80	80	46.7	55	56.7	60	50	66.7	48.3	20
C	55	78.3	76.7	80	46.7	65	63.3	71.7	51.7	41.7
D	76.7	86.7	73.3	78.3	73.3	66.7	66.7	65	50	55
Average	71.3	83.3	67.9	68.3	62.1	66.3	60	67.5	48.3	39.2

Fig. 6. Comparison of confusion matrix with/without cover [%]

5 Conclusion

We proposed a method to improve ultrasound-based gesture recognition by attaching cover to the microphone. Since the proposed method utilizes a 3D printed cover and a pair of a speaker and microphone that is embedded to the device, electrical additional devices are not needed to improve gesture recognition. We proposed and compared designs of microphone cover. Evaluation results confirmed that recognition accuracy increased 12% by using the proposed method.

References

1. Kratz, S., Rohas, M.: HoverFlow: expanding the design space of around-device interaction. In: Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI 2009), no. 4, pp. 1–8 (2009)
2. Manabe, H.: Multi-touch gesture recognition by single photoreflector. In: Proceedings of the 26th ACM Symposium on User Interface Software and Technology (UIST 2013), pp. 15–16 (2013)
3. Nandakumar, R., Iyer, V., Tan, D., Gollakota, S.: FingerIO: using active sonar for fine-grained finger tracking. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2016), pp. 1515–1525 (2016)

4. Ruan, W., Sheng, Q.Z., Yang, L., Gu, T., Xu, P., Shangguan, L.: AudioGest: enabling fine-grained hand gesture detection by decoding echo signal. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2016), pp. 474–485 (2016)
5. Song, J., Sörös, G., Pece, F., Fanello, S.R., Izadi, S., Keskin, C., Hilliges, O.: In-air gestures around unmodified mobile devices. In: Proceedings of the 27th ACM Symposium on User Interface Software and Technology (UIST 2014), pp. 319–329 (2014)